Relation Classification with Entity Type Restriction

Shengfei Lyu,   Huanhuan Chen[*]
School of Computer Science and Technology
University of Science and Technology of China, Hefei, China
saintfe@mail.ustc.edu.cn, hchen@ustc.edu.cn

Abstract

Relation classification aims to predict a relation between two entities in a sentence. The existing methods regard all relations as the candidate relations for the two entities. These methods neglect the restrictions on candidate relations by entity types, which leads to some inappropriate relations being candidate relations. In this paper, we propose a novel paradigm, RElation Classification with ENtity Type restriction (RECENT), which exploits entity types to restrict candidate relations. Specially, the mutual restrictions of relations and entity types are formalized and introduced into relation classification. Besides, the proposed paradigm, RECENT, is model-agnostic. Based on two representative models GCN and SpanBERT respectively, RECENT$_{\text{GCN}}$ and RECENT$_{\text{SpanBERT}}$ are trained in RECENT[1]. Experimental results on a standard dataset indicate that RECENT improves the performance of GCN and SpanBERT by 6.9 and 4.4 F1 points, respectively. Especially, RECENT$_{\text{SpanBERT}}$ achieves a new state-of-the-art on TACRED.

1 Introduction

Relation classification, a supervised version of relation extraction, aims to predict a relation between two entities in a sentence. Relation classification is an important step to construct knowledge bases from a large number of unstructured texts (Trisedya et al., 2019), which benefits many natural language processing applications, such as natural language generation (Kang and Hashimoto, 2020) and question answering (Zhao et al., 2020).

Recently, the majority of methods make use of various neural network architectures to learn a fixed-size representation for a sentence and its entities with various language features, such as

---

[*]Corresponding author.
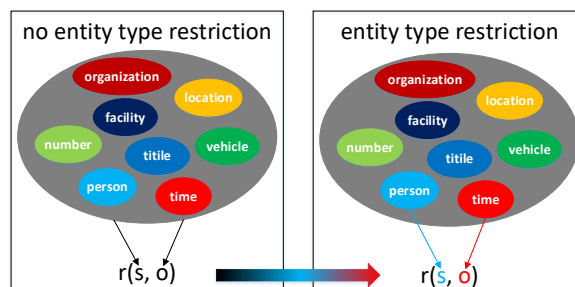[1]Our code is available at https://github.com/Saintfe/RECENT.

Figure 1: A relation restricts entities with appropriate types. In the figure, r is *who-is-born-when*. Different colored ellipses represent entities with different types.
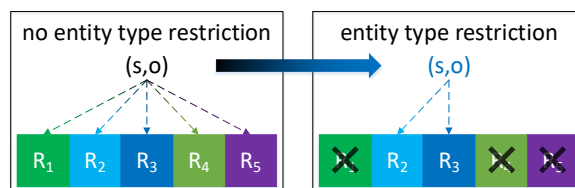

Figure 2: Entity type restriction for relation classification. According to entity type restriction, the number of candidate relations reduces from 5 (left) to 2 (right).

part of speech (POS), entity types, and dependency trees. Dependency trees that are parsed from sentences are exploited by GCN (Kipf and Welling, 2017) to model sentences (Zhang et al., 2018; Guo et al., 2019). As a sequence of words, a sentence is modeled by LSTM (Hochreiter and Schmidhuber, 1997) and its entity positions are involved with the attention mechanism (Zhang et al., 2017). More recently, pretrained language models (Devlin et al., 2019; Baldini Soares et al., 2019; Joshi et al., 2020) achieve good performance in relation classification since they are pretrained on massive corpora.

To recap, these methods utilize an encoder architecture (Badrinarayanan et al., 2017) to obtain a representation for a sentence. In other words, they only focus on the modeling of sentences and
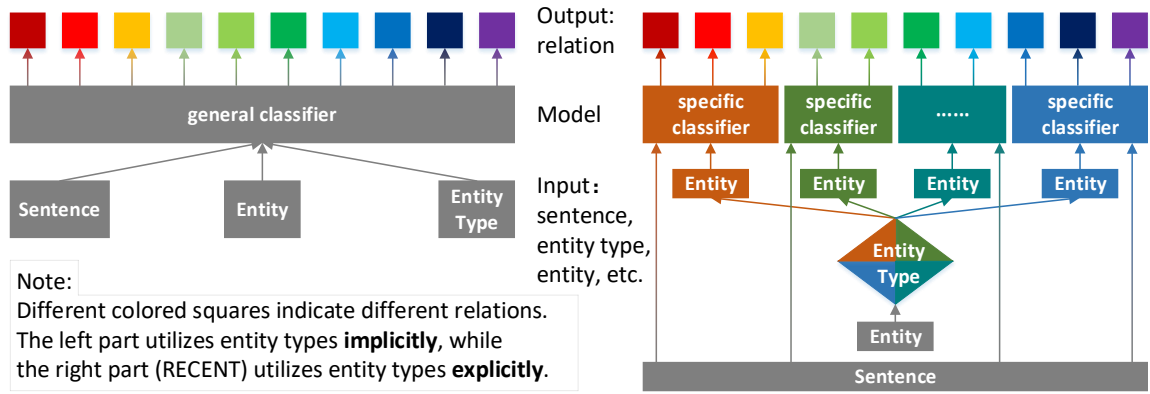
Figure 3: Relation classification with entity type restriction. The left part does not consider the restriction of entity types on relations and only feeds entity types as features into a general classifier. The right part explicitly utilizes entity types to restrict candidate relations and learns a specific classifier for each pair of entity types.

treat relations as labels[2] to be classified. However, in this process, these methods inevitably lose the semantics of relations. Take the mutual restrictions between a relation and entity types as an example. In Figure 1, the relation *who-is-born-when* restricts its first entity to be a person and the second one to be a time. Conversely, entity types can also restrict candidate relations in relation classification. As illustrated in Figure 2, some inappropriate relations can be discarded from candidate relations by entity type restriction. However, the current methods neglect the restriction of entity types on relations so that some inappropriate relations are regarded as candidate relations, which further hurts their performance.

To solve the above problem, a novel paradigm, RElation Classification with ENtity Type restriction (RECENT), is proposed to exploit entity types to restrict candidate relations. As the basis of the paradigm, the mutual restrictions of relations and entity types are formalized. With the entity type restriction, some inappropriate relations are discarded from the candidate relations of a specific pair of entity types, as illustrated in Figure 2. A specific classifier with a specific set of candidate relations is individually learned for each pair of entity types (Figure 3). Therefore, the proposed paradigm, RECENT, can eliminate the interference from inappropriate candidate relations.

The contributions are summarized as follows:

- The mutual restrictions of relations and entity types are formalized.

- A novel paradigm, RECENT, is proposed to

exploit entity types to restrict candidate relations in relation classification.

- A new state-of-the-art is achieved on TACRED.

2 Proposed Paradigm

Before introducing the proposed paradigm RECENT, the mutual restrictions between a relation and a pair of entities are formalized as the basis of RECENT.

2.1 Relation Function

When a binary relation is considered as a function, this relation has two entities as its two arguments. Formally, this relation is formalized as $r(s, o)$, where r denotes the **r**elation and s, o denote the first (**s**ubject) entity and the second (**o**bject) entity, respectively. The range of this relation contains two discrete values $\{0, 1\}$:

$$r(s, o) = \begin{cases} 1 & r \text{ holds between } s \text{ and } o, \\ 0 & \text{otherwise.} \end{cases} \quad (1)$$

In a broad sense, the domain of this relation can be any pair of entities. However, when a pair of entities with inappropriate types is fed into a specific relation, the relation can directly return 0, no need to consider the compositional semantics of the relation and the pair of entities. For example, a specific relation *who-is-born-when* expects the first argument to be a person and the second one to be a time. Therefore, (*apple*, *Steven Jobs*) is a pair of inappropriate entities for this relation so that *who-is-born-when(apple, Steven Jobs)* returns 0 without considering the compositional semantics,

---

[2]Specifically, these meaningful relations are treated as meaningless numbers, such as 0, 1, 2.

since *apple* may refer to either a kind of fruit or a company (not a person) and *Steven Jobs* may refer to a famous person (not a time).

Only when a relation receives a pair of appropriate entities whose types match it, the combination of the relation and the entities might make sense (i.e., the function defined in Eq. 1 may return 1). In this case, it is meaningful to further verify the correctness of the compositional semantics. From this perspective, in a narrow sense, the domain (denoted by D_r) of a relation (r) is defined as follows:

$$D_r = \{(s, o) | ts \in S(r) \text{ and } to \in O(r)\}, \quad (2)$$

where ts and to denote the **t**ypes of the subject entity (s) and the object entity (o), respectively. $S(r)$ and $O(r)$ are the appropriate types of r on the subject entity (s) and the object entity (o), respectively.

2.2 Entity Type Restriction

In the previous subsection, the narrow domain of a relation restricts entities whose types need to match the relation. Conversely, given a pair of entities whose types are known, the candidate relations of the entities are also restricted, since the match between relations and entity types is mutual.

Formally, given a pair of entities (s, o) and their types (ts, to), its candidate relations (denoted by $R_{(ts,to)}$) are restricted into a limited set:

$$\begin{aligned} R_{(ts,to)} &= \{r \in R | (s, o) \in D_r\} \\ &= \{r \in R | ts \in S(r) \text{ and } to \in O(r)\}, \end{aligned} \quad (3)$$

where R denotes all possible relations. When the types (ts, to) of a pair of entities (s, o) are explicitly utilized to restrict its candidate relations, the candidate relations reduce from all possible relations R into a rather smaller set $R_{(ts,to)}$.

2.3 Relation Classification

Unlike traditional methods that classify a sentence and its entities on all candidate relations R (the left part of Figure 3), the proposed paradigm, RECENT learns a specific classifier with smaller and more precise candidate relations for each pair of entity types (the right part of Figure 3), based on entity type restriction in the previous subsection.

The procedure of RECENT is summarized in Algorithm 1. In the learning phase, all sentences are first grouped by types of their entities (line 1). For each group (marked as g) with a specific pair of

entity types (ts,to), the candidate relations $R_{(ts,to)}$ for the group g are obtained by aggregating the relations in the group g (line 3). Then, a specific classifier (marked by f_g) that maps sentences and their entities in g to $R_{(ts,to)}$, is learned for the group g (line 4). In the prediction phase, given a new sample (se, s, o, ts, to), a group (marked as g') is matched by the entity types (ts, to) (line 6). Then, the classifier $f_{g'}$ learned on the group g' is utilized to predict a relation according to the input (se, s, o) (line 7).

From the 4th line of Algorithm 1, the proposed paradigm RECENT is model-agnostic, which means that RECENT is theoretically compatible with many relation classification models.

Algorithm 1 RECENT

Learning Phase:
>   **Input:** $\mathcal{D} = \{(se_i, s_i, o_i, ts_i, to_i, r_i) | i = 1, 2, ..., N\}$ where the subscript i indicates the ith sample, se is **se**ntence, s is **s**ubject entity, o is **o**bject entity, ts is **t**ype of **s**ubject entity, to is **t**ype of **o**bject entity, r is **r**elation.
>   **Output:** Multiple classifiers.

1: Group sentences by entity types.
2: **for** each **g**roup g (enity types (ts, to)) **do**
3:   aggregate relations in the group as candidate relations $R_{(ts,to)}$ defined in Eq. 3.
4:   learn a classifier (marked as f_g) on the group that maps $\{(se_i, s_i, o_i) \in g\}$ to $R_{(ts,to)}$.
5: **end for**

Prediction Phase:
>   **Input:** A new sample $\{se, s, o, ts, to\}$, each specific classifier for each pair of entity types.
>   **Output:** A relation.

6: match the sample to a group (marked as g') according to the entity types (ts, to).
7: Use the classifier ($f_{g'}$) learned on the group to map (se, s, o) to a relation.
8: **return** the relation.

---

3 Experiments

3.1 Dataset

The proposed paradigm RECENT is evaluated on TACRED[3] (Zhang et al., 2017). TACRED contains 41 semantic relations and a special *no_relation* over 106,264 sentences. The subject entities in TACRED are classified into two types: *PERSON* and

---

[3] https://catalog.ldc.upenn.edu/ LDC2018T24

392

ORGANIZATION while the object entities are categorized into 16 fine-grained types, such as *LOCATION* and *TIME*. Namely, entity types are known. By convention, the micro-averaged F1 score (abbreviated as F1) is reported on TACRED.

3.2 Experimental Setup

Since *no_relation* is a candidate relation of each pair of entity types in TACRED, a binary classifier is first learned to distinguish between 41 semantic relations and *no_relation*. In this way, each pair of entity types reduces one candidate relation (i.e. *no_relation*) in RECENT. If the binary classifier predicts *no_relation* for a pair of entities, then the final relation for them is *no_relation*. Otherwise, their specific semantic relation is further predicted in RECENT.

Base Models The proposed paradigm RECENT is model-agnostic. Two representative models that are GCN (Zhang et al., 2018) and SpanBERT (Joshi et al., 2020) are selected as base models (line 4 in Algorithm 1). For a fair comparison with a base model, all classifiers (including the binary classifier) in RECENT are trained by the base model. The corresponding models in the paper are denoted as RECENT$_{GCN}$ and RECENT$_{SpanBERT}$.

Hyperparameters For RECENT$_{GCN}$, the path-centric pruning K is set to 1 as GCN (Zhang et al., 2018). The learning rates for all classifiers in RECENT$_{GCN}$ are set to 0.3. For RECENT$_{SpanBERT}$, the learning rates for all classifiers are chosen from {5e-6, 1e-5, 2e-5, 3e-5, 5e-5} as SpanBERT.

Compared Models Extensive models in relation classification are regarded as comparison models. They include PA-LSTM (Zhang et al., 2017), C-GCN (Zhang et al., 2018), AGGCN (Guo et al., 2019), C-AGGCN (Guo et al., 2019), MTB (Baldini Soares et al., 2019), KnowBert (Peters et al., 2019), SpanBERT-ALT (Lyu et al., 2020), KEPLER (Wang et al., 2020b), K-Adapter (Wang et al., 2020a), and LUKE (Yamada et al., 2020). To save space, please refer to the original papers of these models for details.

3.3 Experimental Results

The experimental results are presented in Table 1. RECENT$_{GCN}$ achieves a significant performance increase on the F1 score above its base model GCN. The absolute increase reaches 6.9 from 64.0 to 70.9.

Model	P	R	F1
PA-LSTM † (Zhang et al., 2017)	65.7	64.5	65.1
C-GCN † (Zhang et al., 2018)	69.9	63.3	66.4
AGGCN † (Guo et al., 2019)	69.9	60.9	65.1
C-AGGCN † (Guo et al., 2019)	71.8	66.4	69.0
GCN † (Zhang et al., 2018)	69.8	59.0	64.0
RECENT$_{GCN}$ (ours)	88.3	59.3	70.9
SpanBERT-ALT † (Lyu et al., 2020)	69.0	73.0	70.9
MTB † (Baldini Soares et al., 2019)	-	-	71.5
KnowBert † (Peters et al., 2019)	71.6	71.4	71.5
KEPLER †* (Wang et al., 2020b)	71.5	72.5	72.0
K-Adapter †* (Wang et al., 2020a)	70.14	74.04	72.04
LUKE † (Yamada et al., 2020)	70.4	**75.1**	72.7
SpanBERT † (Joshi et al., 2020)	70.8	70.9	70.8
RECENT$_{SpanBERT}$ (ours)	**90.9**	64.2	**75.2**

Table 1: Results on the TACRED dataset. P and R indicate precision and recall, respectively. Bold marks the highest values among models. † marks results reported in the original papers. * marks results from preprint papers.

The main contribution for the F1 increase is the improved precision that greatly increases from 69.8 to 88.3. The great increase in precision, which might result from the restriction on candidate relations by entity types in RECENT, indicates the effectiveness of the proposed paradigm RECENT. Besides, RECENT$_{GCN}$ suppresses the compared models that do not include pretrained language models.

Similarly, RECENT$_{SpanBERT}$ overtakes its base model SpanBERT by absolute 4.4 points on F1. The great soar (absolute 20.1 points) on precision contributes the superior F1 of RECENT$_{SpanBERT}$. Unfortunately, the decline in recall limits the further improvement of F1. This might be due to sample imbalance of candidate relations, which will be further studied in future work. On the whole (i.e. F1), RECENT$_{SpanBERT}$ outperforms all the compared models. Especially, RECENT$_{SpanBERT}$ exceeds the state-of-the-art LUKE model[4] by 2.5 F1 points and achieves a new state-of-the-art.

3.4 Error Analysis of GCN

This subsection analyzes the influence of a baseline model (i.e. GCN) that neglects the restriction of entity types on relations. We retrain a GCN model and the model achieves 68.4 precision, 60.2 recall, and 64.1 F1 (Table 2), which are similar to the

---

[4]LUKE achieves the state-of-the-art (72.7) on the published papers. Cohen et al. (2020) report a new state-of-the-art (74.8) in the preprint way. Anyway, RECENT$_{SpanBERT}$ achieves a new state-of-the-art (75.2).

Model	P	R	F1	FP	FP(ET)
GCN	68.4	60.2	64.1	1,323	144

Table 2: Results of our trained GCN on the TACRED dataset. P and R indicate precision and recall, respectively. FP indicates the number of false positives and FP(ET) indicates the number of false positives that break the entity type restriction.

results in its reported paper (Zhang et al., 2018). Observing the prediction results of the model, we find that 1) 1,323 examples are false positives in the test set of TACRED, 2) 144 (about 11%) false positives among them break the entity type restriction. Namely, GCN can make about 89% of false positives meet the entity type restriction, by implicitly using entity types. However, about 11% of false positives still break the restriction. The false positives broken down by relations are counted in Appendix A. In details, false positives broken down by relations are weakly negatively correlated with the amount of training data of relations, where the correlation coefficient is -0.39. This infers that fewer training examples of relations may lead to more false positives of relations.

4 Conclusion

In the paper, a novel paradigm, RECENT, is proposed by entity type restriction. RECENT reduces candidate relations for each pair of entity types by the mutual restrictions between relations and entity types. RECENT is model-agnostic. $RECENT_{GCN}$ and $RECENT_{SpanBERT}$ that are based on two representative models GCN and SpanBERT respectively, outperform their counterparts on the standard dataset TACRED, which empirically indicates the effectiveness of the proposed paradigm RECENT. Especially, $RECENT_{SpanBERT}$ achieves a new state-of-the-art on TACRED.

Acknowledgments

References

Vijay Badrinarayanan, Alex Kendall, and Roberto Cipolla. 2017. SegNet: A deep convolutional encoder-decoder architecture for image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39(12):2481–2495.

Livio Baldini Soares, Nicholas FitzGerald, Jeffrey Ling, and Tom Kwiatkowski. 2019. Matching the blanks: Distributional similarity for relation learning. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 2895–2905, Florence, Italy. Association for Computational Linguistics.

Amir D. N. Cohen, Shachar Rosenman, and Yoav Goldberg. 2020. Relation extraction as two-way span-prediction. *CoRR*, abs/2010.04829.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. BERT: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 4171–4186, Minneapolis, Minnesota. Association for Computational Linguistics.

Zhijiang Guo, Yan Zhang, and Wei Lu. 2019. Attention guided graph convolutional networks for relation extraction. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 241–251, Florence, Italy. Association for Computational Linguistics.

Sepp Hochreiter and Jürgen Schmidhuber. 1997. Long short-term memory. *Neural computation*, 9(8):1735–1780.

Mandar Joshi, Danqi Chen, Yinhan Liu, Daniel S. Weld, Luke Zettlemoyer, and Omer Levy. 2020. SpanBERT: Improving pre-training by representing and predicting spans. *Transactions of the Association for Computational Linguistics*, 8:64–77.

Daniel Kang and Tatsunori Hashimoto. 2020. Improved natural language generation via loss truncation. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 718–731, Online. Association for Computational Linguistics.

Thomas N. Kipf and Max Welling. 2017. Semi-supervised classification with graph convolutional networks. In *Proceedings of the 5th International Conference on Learning Representations, Toulon, France*.

S. Lyu, J. Cheng, X. Wu, L. Cui, H. Chen, and C. Miao. 2020. Auxiliary learning for relation extraction. *IEEE Transactions on Emerging Topics in Computational Intelligence*, pages 1–10.

Matthew E. Peters, Mark Neumann, Robert Logan, Roy Schwartz, Vidur Joshi, Sameer Singh, and Noah A.

Smith. 2019. Knowledge enhanced contextual word representations. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 43–54, Hong Kong, China. Association for Computational Linguistics.

Bayu Distiawan Trisedya, Gerhard Weikum, Jianzhong Qi, and Rui Zhang. 2019. Neural relation extraction for knowledge base enrichment. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 229–240, Florence, Italy. Association for Computational Linguistics.

Ruize Wang, Duyu Tang, Nan Duan, Zhongyu Wei, Xuanjing Huang, Jianshu Ji, Guihong Cao, Daxin Jiang, and Ming Zhou. 2020a. K-Adapter: Infusing knowledge into pre-trained models with adapters. *CoRR*, abs/2002.01808v5.

Xiaozhi Wang, Tianyu Gao, Zhaocheng Zhu, Zhiyuan Liu, Juanzi Li, and Jian Tang. 2020b. KEPLER: A unified model for knowledge embedding and pre-trained language representation. *CoRR*, abs/1911.06136v3.

Ikuya Yamada, Akari Asai, Hiroyuki Shindo, Hideaki Takeda, and Yuji Matsumoto. 2020. LUKE: Deep contextualized entity representations with entity-aware self-attention. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 6442–6454, Online. Association for Computational Linguistics.

Yuhao Zhang, Peng Qi, and Christopher D. Manning. 2018. Graph convolution over pruned dependency trees improves relation extraction. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 2205–2215, Brussels, Belgium. Association for Computational Linguistics.

Yuhao Zhang, Victor Zhong, Danqi Chen, Gabor Angeli, and Christopher D. Manning. 2017. Position-aware attention and supervised data improve slot filling. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, pages 35–45, Copenhagen, Denmark. Association for Computational Linguistics.

Xinyan Zhao, Feng Xiao, Haoming Zhong, Jun Yao, and Huanhuan Chen. 2020. Condition aware and revise transformer for question answering. In *Proceedings of The Web Conference*, page 2377–2387.

A The Statistics of False Positives

Table 3 presents false positives broken down by relations of our trained GCN on the TACRED dataset. In details, false positives broken down by relations are weakly negatively correlated with the amount of training data of relations, where the correlation coefficient is -0.39.

Relation	FP
org:alternate_names	53
org:city_of_headquarters	30
org:country_of_headquarters	70
org:dissolved	2
org:founded	9
org:founded_by	46
org:member_of	18
org:members	31
org:number_of_employees/members	9
org:parents	62
org:political/religious_affiliation	4
org:shareholders	10
org:stateorprovince_of_headquarters	16
org:subsidiaries	37
org:top_members/employees	61
org:website	1
per:age	13
per:alternate_names	11
per:cause_of_death	37
per:charges	35
per:children	31
per:cities_of_residence	90
per:city_of_birth	3
per:city_of_death	19
per:countries_of_residence	93
per:country_of_birth	5
per:country_of_death	9
per:date_of_birth	3
per:date_of_death	36
per:employee_of	98
per:origin	46
per:other_family	60
per:parents	49
per:religion	16
per:schools_attended	13
per:siblings	25
per:spouse	24
per:stateorprovince_of_birth	4
per:stateorprovince_of_death	10
per:stateorprovinces_of_residence	40
per:title	94

Table 3: False positives (FP) broken down by relations of our trained GCN on the TACRED dataset.